Chapter 7

Rough Sets Theory

—Rough sets theory (RST) as a mathematical methodology for data analysis were introduced by Pawlak (Pawlak, 1991). They provide a powerful tool for data analysis and knowledge discovery from imprecision and uncertain data.

—The rough sets methodology is based on the premise that lowering the degree of precision in the data makes the data pattern more visible.

—The rough sets approach can be considered as a formal framework for discovering patterns from imperfect data. The results of the rough approach are presented in the form of classification and decision rules derived from given data sets.

—RST operates may be as a knowledge representation system or information system. An information system (S) consisting of four parts is shown as:

$$S = (U, A, V, f),$$

where

U is a non-empty set of objects;

A is the collection of objects; we have $A = C \bigcup D$ and $C \bigcap D = \phi$, where C is a non-empty set of condition attributes, and D is a non-empty set of decision attributes;

V is the union of attribute domains, i.e., $V = \underset{a \in A}{U} V_a$, where V_a is a finite attribute domain and the elements of V_a are called values of attribute a;

f is an information function such that $f(u_i, a) \in V_a$ for every $a \in A$ and

$u_i \in U$.

— Every object that belongs to *U* is associated with a set of values corresponding to the condition attributes *C*, and the decision attributes *D*.

— If *R* is an equivalence relation over *U*, then U/R we mean the family of all equivalence classes of *R* (or classification of *U*) referred to as categories or concepts of *R*, and $[x]_R$ denoted a category in *R* containing an element $x \in U$.

— For every set of attributes $B \subseteq A$, an indiscernibility relation $Ind(B)$ is defined as: two objects x_i and x_j are indiscernible by the set of attributes B in A, if $b(x_i) = b(x_j)$, for every $b \subset B$.

Example 1, Consider a data set containing the results of three measurements $\{a_1, a_2, a_3\}$ performed for 10 objects.

U	a_1	a_2	a_3
x_1	2	1	3
x_2	3	2	1
x_3	2	1	3
x_4	2	2	3
x_5	1	1	4
x_6	1	1	2
x_7	3	2	1
x_8	1	1	4
x_9	2	1	3
x_{10}	3	2	1

$U = \{x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9, x_{10}\}$
$A = \{a_1, a_2, a_3\}$

$V_{a_1} = \{1,2,3\}$, $V_{a_2} = \{1,2\}$, $V_{a_3} = \{1,2,3,4\}$.

U / B:

$$C_1 = \{x_1, x_3, x_9\}$$
$$C_2 = \{x_2, x_7, x_{10}\}$$
$$C_3 = \{x_4\}$$
$$C_4 = \{x_5, x_8\}$$
$$C_5 = \{x_6\}$$

Lower approximation and upper approximation

— The rough sets approach to data analysis hinges on two basic concepts, namely the *lower approximation* and *upper approximation* of a data set. The lower and upper approximations can also be presented in an equivalent form as shown as:

The lower approximation of the set $X \subseteq U$ and $B \subseteq A$:

$$\underline{B}(X) = \{x_i \in U : [x_i]_R \subset X\}$$

The upper approximation of the set $X \subseteq U$ and $B \subseteq A$:

$$\overline{B}(X) = \{x_i \in U : [x_i]_R \cap X \neq \phi\}$$

Example 2, Let us assume that we are interested in the subset X of five objects $\{x_1, x_3, x_4, x_5, x_9\}$.

By example 1, we have

$$C_1 = \{x_1, x_3, x_9\}$$
$$C_2 = \{x_2, x_7, x_{10}\}$$
$$C_3 = \{x_4\}$$
$$C_4 = \{x_5, x_8\}$$
$$C_5 = \{x_6\}$$

Thus,

$$\underline{B}(X) = C_1 \cup C_3 = \{x_1, x_3, x_9, x_4\}$$

$$\overline{B}(X) = C_1 \cup C_3 \cup C_4 = \{x_1, x_3, x_9, x_4, x_5, x_8\}$$

Accuracy of approximation

— An accuracy measure of a set X in $B \subseteq A$ is defined as:

$$\mu_B(X) = \frac{card(\underline{B}(X))}{card(\overline{B}(X))}$$

Quality of classification

— The quality of classification is defined as:

$$\gamma_B(X) = \frac{card(\underline{B}(X))}{card(U)}$$

Core and reducts of attributes values

— If the set of attributes is dependent, then we interested in finding all possible minimal subsets of the attributes, which leads to the same number of elementary sets as the whole attributes. Such minimal subsets $S(S \subseteq A)$ are called *reducts*, denoted as $RED(S)$, the $core = \bigcap RED(S)$.

— A *reduct* of the set of condition attributes F ($F \subseteq C$) with respect to a set of decision attributes D is a subset $RED(F, D)$ of C, which satisfies the following two criteria [1]:

(1) $\gamma(C, D) = \gamma(RED(F, D), D)$;

(2) No attributes can be eliminated from $RED(F, D)$ without effecting the requirement (1).

Discernibility Matrix

— Let $S=(U,A,V,f)$ be a information system with $U=\{x_1,x_2,x_3,\ldots,x_n\}$. By an discernibility matrix of *S*, denoted $M(S)$, we will mean *nxn* matrix defined as:

$$(c_{ij})=\{a\in A \mid a(x_i)\neq a(x_j)\},\quad for\ \ i,j=1,2,\ldots,n\}$$

Thus entry c_{ij} is the set of all attributes which discern objects x_i and x_j.

Discernibility Function

— The discernibility matrix can be used to find the minimal subset of attributes, which leads to the partition of the data as whole set of attributes. To do this, we have to construct the discernibility function $f(A)$.

— The $f(A)$ is a Boolean function constructed in following way:

- Each attribute from the set of attributes, which discern two elementary sets, (a_1,a_2,a_3), the Boolean function attains the form $(a_1+a_2+a_3)$, or it can be present as $(a_1\vee a_2\vee a_3)$.
- If the set of attribute is empty, we assign to it the Boolean constant 1.

Example 3 (unsupervised)

U	a_1	a_2	a_3
x_1	2	1	3
x_2	3	2	1
x_3	2	1	3
x_4	2	2	3
x_5	1	1	4
x_6	1	1	2
x_7	3	2	1
x_8	1	1	4
x_9	2	1	3
x_{10}	3	2	1

U / B

$C_1 = \{x_1, x_3, x_9\};$
$C_2 = \{x_2, x_7, x_{10}\};$
$C_3 = \{x_4\}$
$C_4 = \{x_5, x_8\};$
$C_5 = \{x_6\}$

then, the idscernibility matrix has shown as:

—									
a_1, a_2, a_3	—								
—	a_1, a_2, a_3	—							
a_2	a_1, a_3	a_2	—						
a_1, a_3	a_1, a_2, a_3	a_1, a_3	a_1, a_2, a_3	—					
a_1, a_3	a_1, a_2, a_3	a_1, a_3	a_1, a_2, a_3	a_3	—				
a_1, a_2, a_3	—	a_1, a_2, a_3	a_1, a_3	a_1, a_2, a_3	a_1, a_2, a_3	—			
a_1, a_3	a_1, a_2, a_3	a_1, a_3	a_1, a_2, a_3	—	a_3	a_1, a_2, a_3	—		
—	a_1, a_2, a_3	—	a_2	a_1, a_3	a_1, a_3	a_1, a_2, a_3	a_1, a_3	—	
a_1, a_2, a_3	—	a_1, a_2, a_3	a_1, a_3	a_1, a_2, a_3	a_1, a_2, a_3	—	a_1, a_2, a_3	a_1, a_2, a_3	—

Example 4 (supervised)

U	a_1	a_2	a_3	D
x_1	2	1	3	1
x_2	3	2	1	2
x_3	2	1	3	1
x_4	2	2	3	2
x_5	1	1	4	3
x_6	1	1	2	3
x_7	3	2	1	2
x_8	1	1	4	3

x_9	2	1	3	1
x_{10}	3	2	1	2

The equivalence classes

$D_1 = \{x_1, x_3, x_9\}$;

$D_2 = \{x_2, x_4, x_7, x_{10}\}$;

$D_3 = \{x_5, x_6, x_8\}$

The discernibi.ity matrix was shown as:

—									
a_1, a_2, a_3	—								
—	a_1, a_2, a_3	—							
a_2	—	a_2	—						
a_1, a_3	a_1, a_2, a_3	a_1, a_3	a_1, a_2, a_3	—					
a_1, a_3	a_1, a_2, a_3	a_1, a_3	a_1, a_2, a_3	—	—				
a_1, a_2, a_3	—	a_1, a_2, a_3	—	a_1, a_2, a_3	a_1, a_2, a_3	—			
a_1, a_3	a_1, a_2, a_3	a_1, a_3	a_1, a_2, a_3	—	—	a_1, a_2, a_3	—		
—	a_1, a_2, a_3	—	a_2	a_1, a_3	a_1, a_3	a_1, a_2, a_3	a_1, a_3	—	
a_1, a_2, a_3	—	a_1, a_2, a_3	—	a_1, a_2, a_3	a_1, a_2, a_3	—	a_1, a_2, a_3	a_1, a_2, a_3	—

Using the discernin ility function $f(A)$ find the minimal subsets, the $f(A)$ has the following form:

$$f(A) = (a_1 + a_2 + a_3)a_2(a_1 + a_3)$$

$$\times(a_1+a_2+a_3)$$
$$\times a_2(a_1+a_3)(a_1+a_2+a_3)$$
$$\times(a_1+a_2+a_3)a_2$$
$$\times(a_1+a_2+a_3)(a_1+a_3)$$
$$\times(a_1+a_2+a_3)(a_1+a_3)$$
$$\times(a_1+a_2+a_3)$$
$$\times(a_1+a_3)(a_1+a_2+a_3)$$
$$\times(a_1+a_2+a_3)$$
$$=a_1a_2+a_2a_3$$

Hence, we have the reducts $\{a_1,a_2\}$ and $\{a_2,a_3\}$, then the core is a_2.

D-information system

U	a_1	a_2	D
x_1	2	1	1
x_2	3	2	2
x_3	2	1	1
x_4	2	2	2
x_5	1	1	3
x_6	1	1	3
x_7	3	2	2
x_8	1	1	3
x_9	2	1	1
x_{10}	3	2	2

D-discernibility matrix

—	a_1,a_2	—	a_2	a_1	a_1	a_1,a_2	a_1	—	a_1,a_2
a_1,a_2	—	a_1,a_2	—	a_1,a_2	a_1,a_2	—	a_1,a_2	a_1,a_2	—
—	a_1,a_2	—	a_2	a_1	a_1	a_1,a_2	a_1	—	a_1,a_2
a_2	—	a_2	—	a_1,a_2	a_1,a_2	—	a_1,a_2	a_2	—
a_1	a_1,a_2	a_1	a_1,a_2	—	—	a_1,a_2	—	a_1	a_1,a_2
a_1	a_1,a_2	a_1	a_1,a_2	—	—	a_1,a_2	—	a_1	a_1,a_2

a_1, a_2	—	a_1, a_2	—	a_1, a_2	a_1, a_2	—	a_1, a_2	a_1, a_2	—
a_1	a_1, a_2	a_1	a_1, a_2	—	—	a_1, a_2	—	a_1	a_1, a_2
—	a_1, a_2	—	a_2	a_1	a_1	a_1, a_2	a_1	—	a_1, a_2
a_1, a_2	—	a_1, a_2	—	a_1, a_2	a_1, a_2	—	a_1, a_2	a_1, a_2	—

D-discernibility function

$f_1(D) = (a_1 + a_2) \times a_2 \times a_1 = (a_1 a_2 + a_2) \times a_1 = a_2 \times a_1 = a_1 a_2$

$f_2(D) = a_1 + a_2$

$f_3(D) = (a_1 + a_2) \times a_2 \times a_1) = a_1 a_2$

$f_4(D) = a_2$

$f_5(D) = a_1$

$f_6(D) = a_1$

$f_7(D) = a_1 + a_2$

$f_8(D) = a_1$

$f_9(D) = a_1 a_2$

$f_{10}(D) = a_1 + a_2$

Final D-information system was shown as:

U	a_1	a_2	D
x_1	2	1	1
x_2	*	2	2
x_3	2	1	1
x_4	*	2	2
x_5	1	*	3
x_6	1	*	3
x_7	*	2	2
x_8	1	*	3
x_9	2	1	1
x_{10}	*	2	2

Extraction rules

1. If $a_1 = 2$ and $a_2 = 1$, then $D = 1$. (3/3)

2. If $a_2 = 2$, then $D = 2$. (4/4)

3. If $a_1 = 1$, then $D = 3$. (3/3)